

**Supplementary information for Morin et al:** Somatic mutation of EZH2 (Y641) in Follicular and Diffuse Large B-cell Lymphomas of Germinal Center Origin

**Contents:**

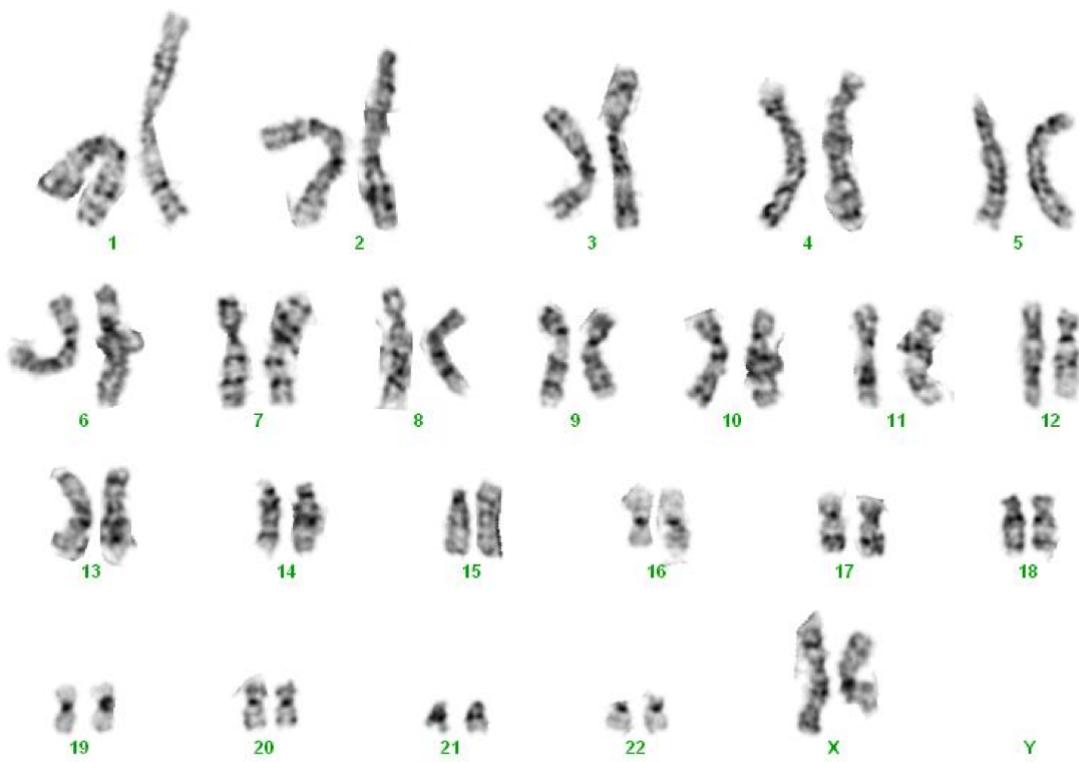
**Supplementary Figures**

- Supplementary Figure 1: Karyotype of “FL patient A”.
- Supplementary Figure 2: aCGH result for FL patient A using flow-sorted tumor DNA and constitutional DNA from peripheral blood as a reference.
- Supplementary Figure 3: Absence of the t(14;18) by FISH result for FL patient A using the LSI *IGH/BCL2* dual color, dual fusion translocation probe.
- Supplementary Figure 4: Overview of Affymetrix copy number analysis result using flow-sorted tumor DNA and constitutional DNA from peripheral blood as a reference.
- Supplementary Figure 5: Overview of copy number alterations identified using WGSS.
- Supplementary Figure 6: Summary of samples analyzed and sequencing method.
- Supplementary Figure 7: Modeling the position of the H3 tail peptide in the EZH2 model.

**Supplementary Tables**

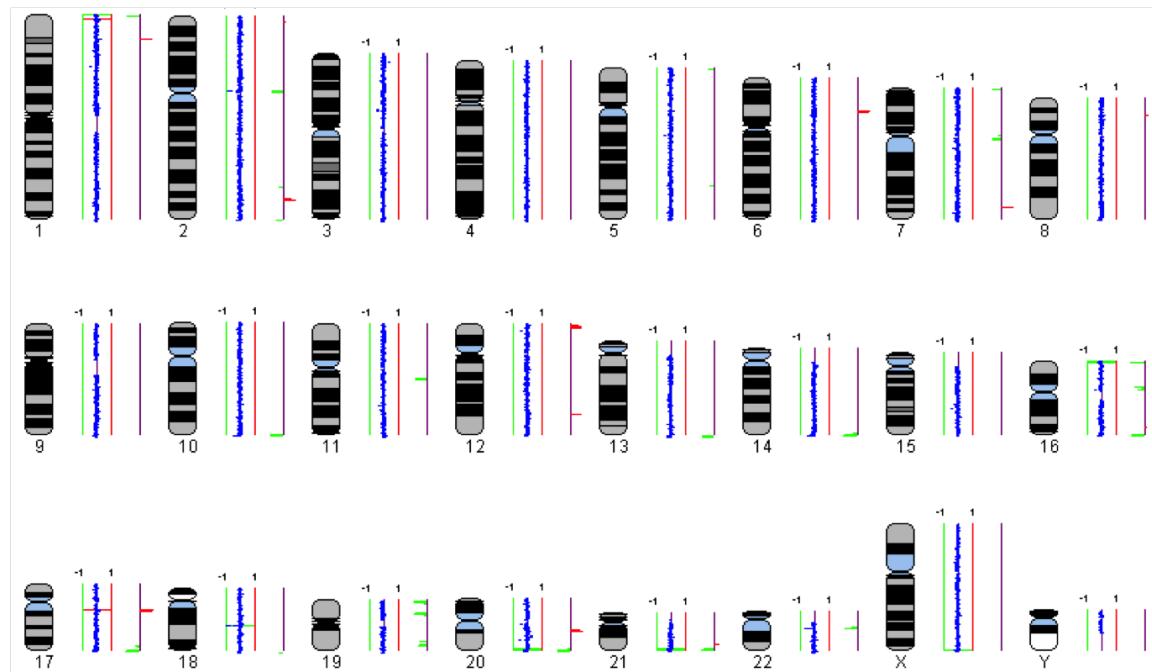
- Supplementary Table 1: Copy number alterations identified by Affymetrix 500k SNP array analysis comparing post-treatment tumor DNA to peripheral blood.
- Supplementary Table 2: Segments identified as copy-number gains or losses in the genome of FL patient A comparing pre-treatment tumor DNA vs peripheral blood WGSS reads.
- Supplementary Table 3: Novel coding SNVs (candidate mutations) identified in WTSS.
- Supplementary Table 4: Coverage of *EZH2* and exon 15 in all RNA-seq libraries.
- Supplementary Table 5: All Y641 mutants detected by Sanger sequencing in FL and DLBCL.
- Supplementary Table 6: Ultra deep targeted re-sequencing of normal B-cell populations and non-GCB lymphomas.
- Supplementary Table 7: Primer sequences

**Supplementary Figure 1:** Karyotype of “FL patient A”.



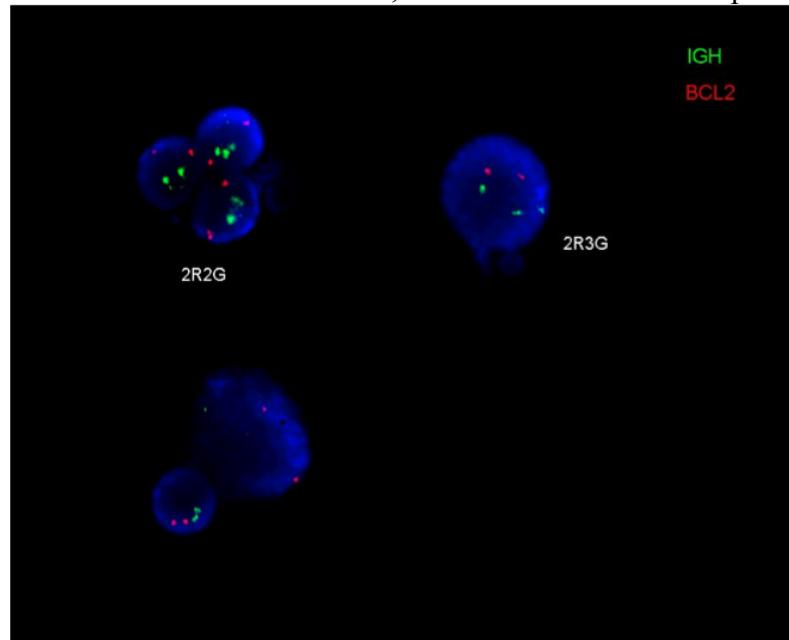
A karyotype produced from the lymph node biopsy from FL patient A shows a normal female chromosomal complement (46, XX). No visual abnormalities were identified.

**Supplementary Figure 2:** aCGH result for FL patient A comparing flow-sorted tumor DNA to constitutional DNA from peripheral blood.

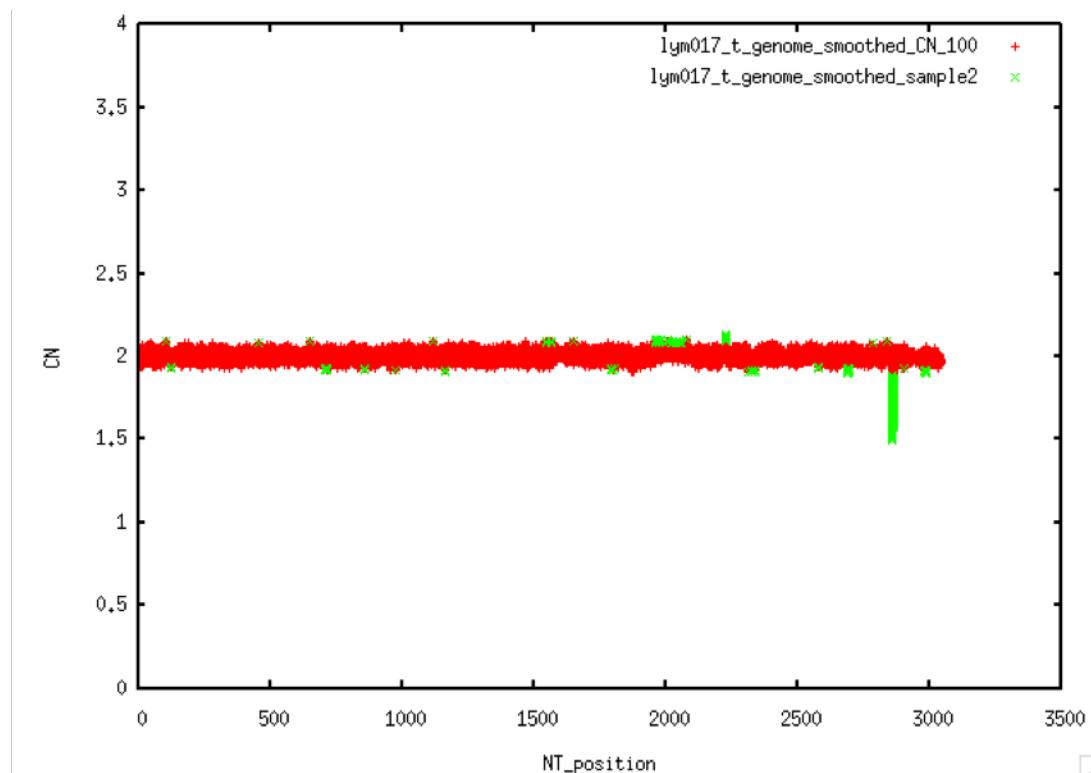


Shown is the karyogram representation of the BAC array CGH data using the “seeGH” software. Losses are depicted as deviations in left (green) and gains as deviations to the right (red). None of the gains or losses represented here were deemed statistically significant by the software.

**Supplementary Figure 3:** Absence of the t(14;18) by FISH result for FL patient A using the LSI *IGH/BCL2* dual color, dual fusion translocation probe.

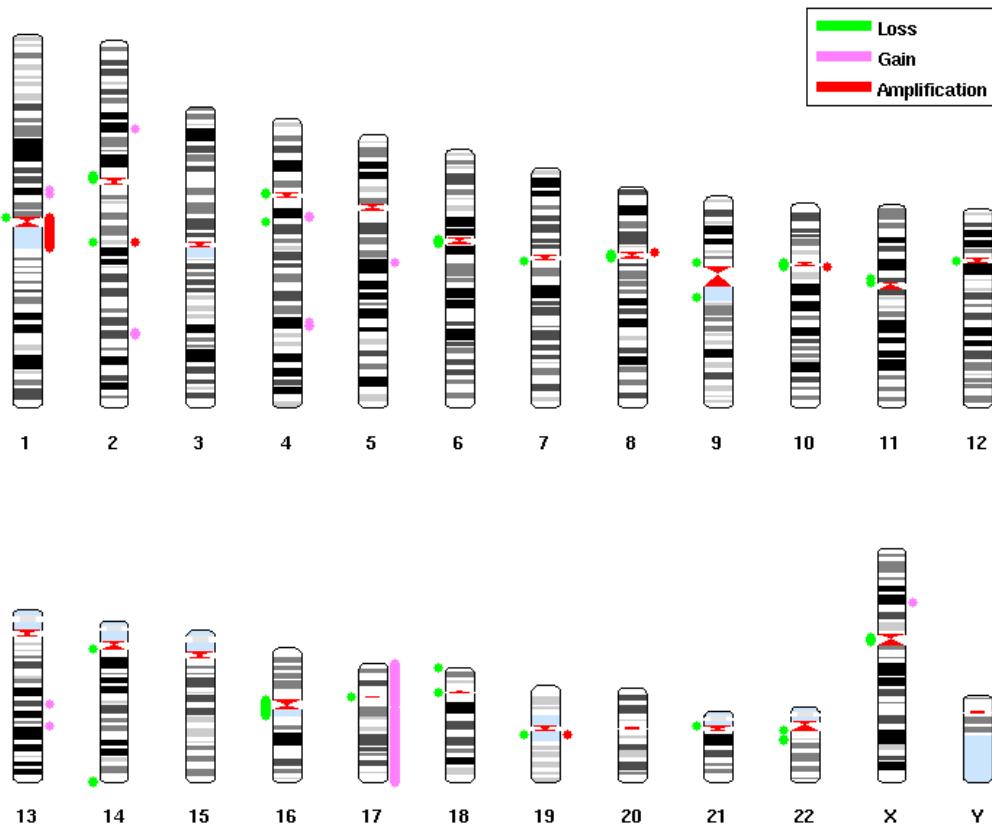


**Supplementary Figure 4:** Overview of Affymetrix copy number analysis result comparing flow-sorted tumor DNA and constitutional DNA to peripheral blood.



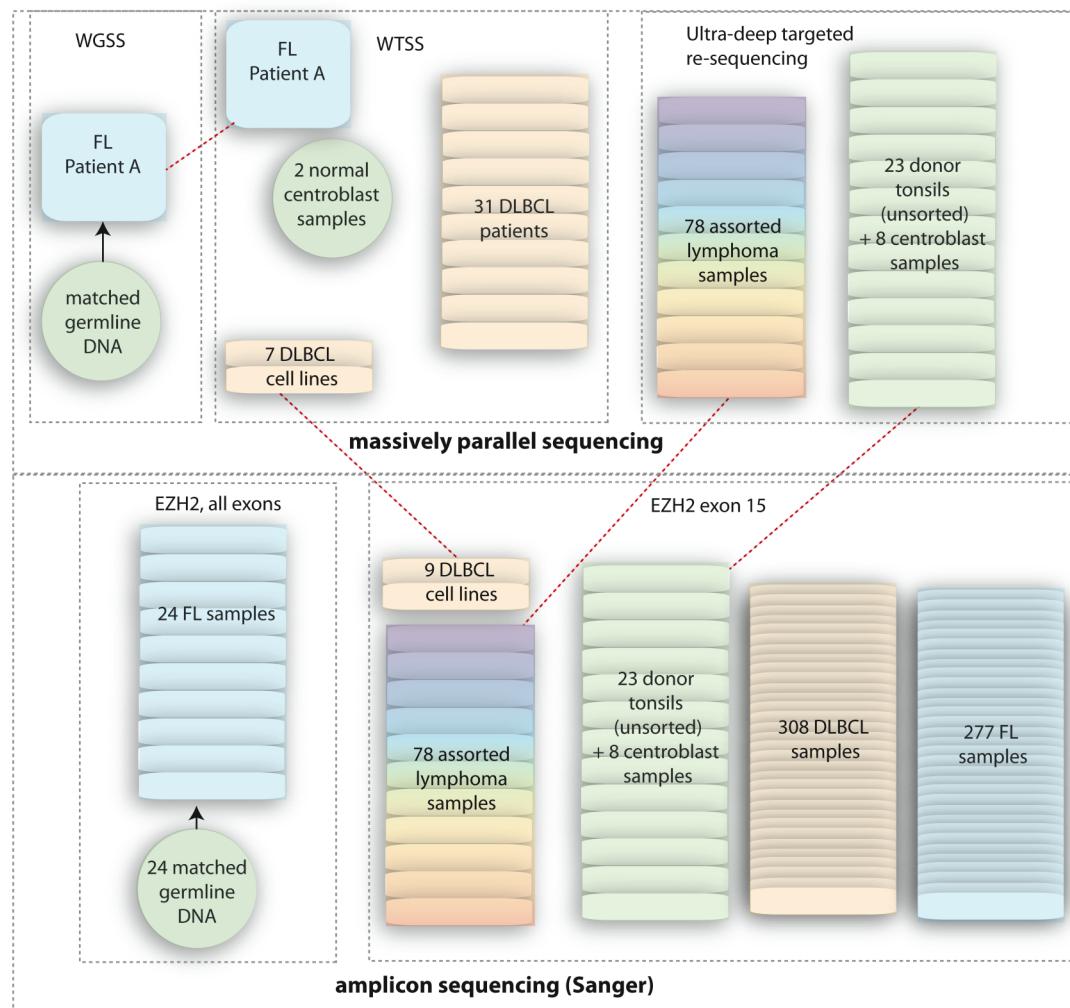
Affymetrix 500k SNP arrays were used to assess the copy number state of a sample from FL patient A taken after treatment failure (in contrast to the sample used for sequencing, which was obtained at diagnosis). Deviations below the horizontal axis reflect deletions/losses and deviations above reflect gains/amplifications. This sample showed minimal significant copy number changes (summarized in Supplementary Table 2).

**Supplementary Figure 5:** Overview of copy number alterations identified using WGSS.



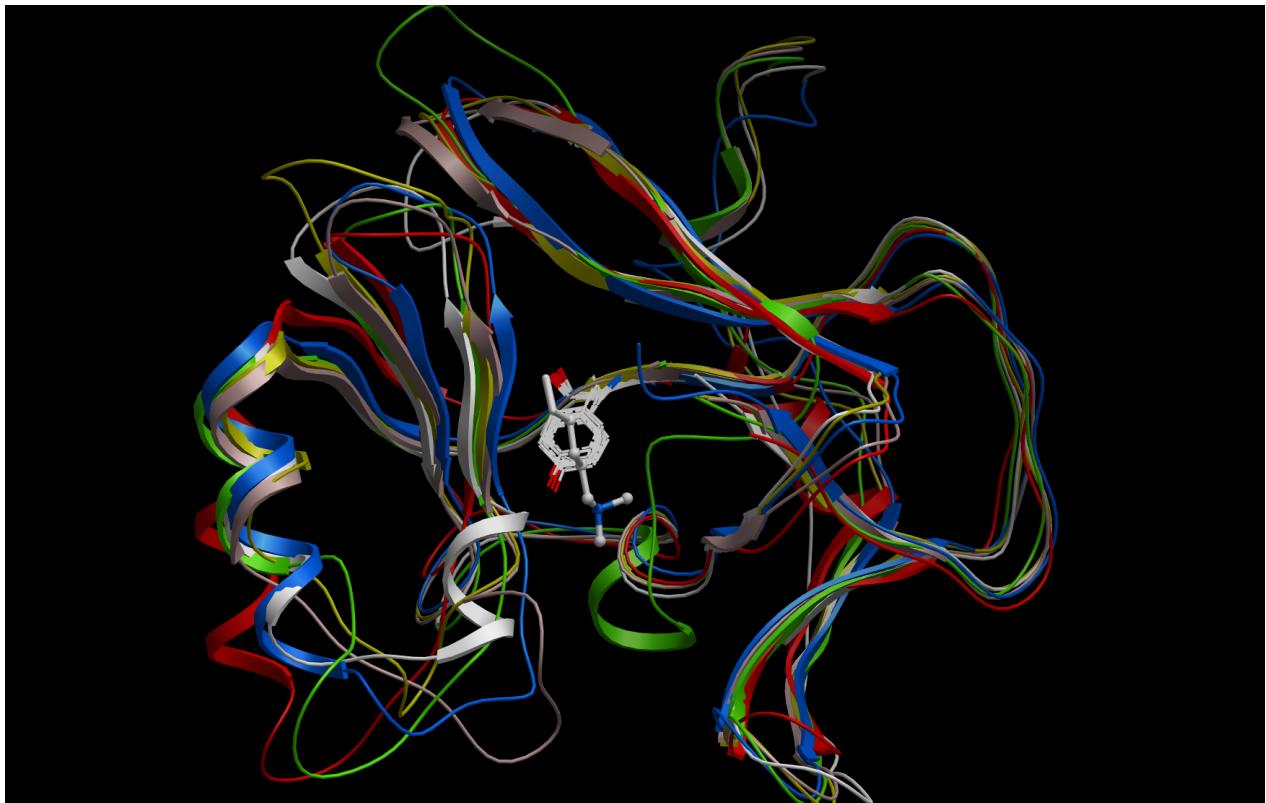
A karyogram representation of FL Patient A tumor vs germline DNA is shown. The genomic sequence reads (Table 1) were processed to identify copy number alterations as described (Methods). Regions are color-coded to indicate the state assigned to them by the Hidden Markov Model. State 1 (deletion) is green, state 2 (normal) is not shown, and states 3 and 4 are pink and red, respectively. No bins were assigned the fifth/high-level amplicon state. Artificial gain and loss states near centromeres are often observed as a technical artifact due to our decreased ability to confidently map reads to these regions. The boundaries of each copy number alteration denoted here are listed in Supplementary Table 3 (below).

**Supplementary Figure 6:** Summary of samples analyzed and sequencing method.



The samples used for initial discovery and confirmation/quantification of the *EZH2* mutations are summarized. Samples sequenced by multiple approaches are connected with a dashed red line. The original observation was identified in both the genome and transcriptome of FL patient A. This mutation was notable as it was not observed in the matched germline DNA of this patient nor was it visible in the transcriptomes of normal centroblasts cells (green). The transcriptomes of 31 DLBCL patient samples were sequenced and revealed that *EZH2* is recurrently mutated in lymphomas. The entire exonic sequence of *EZH2* was sequenced from tumor and germline DNA from 24 FL patients by PCR/capillary sequencing. The mutated exon was also sequenced by capillary sequencing in 308 DLBCL patient samples and 9 cell lines as well as 277 FL patient samples (including those sequenced by WTSS). Additional non-GCB lymphoma samples include 23 mantle-cell lymphomas, 25 peripheral T-cell lymphomas and 30 small lymphocytic lymphomas. None of the additional samples contained mutations in exon 15 of *EZH2*. To confirm that *EZH2* mutations were not “missed” by capillary sequencing, we also performed ultra-deep targeted re-sequencing on all additional samples (tonsil and other non-GCB lymphomas) using Illumina massively parallel sequencing (Methods).

**Supplementary Figure 7:** Modeling the position of the H3 tail peptide in the EZH2 model.



The model constructed based on MLL1 (PDB 2W5Z) is colored red. Models based on other structures are colored as following: EHMT1 (2RFI, yellow); DIM-5 (1PEG, green); SUV39H2 (2R3A, grey); SETD2 (3H6L, blue); EHMT2 (2O8J, pink). All models show Y641 (thin wire, light grey) in a similar conformation, indicating that either model could accurately represent the positioning of Y641 in the SET domain. The K4 residue from MLL1 and K9 residues from EHMT1 and Dim-5 are shown in stick models.

## Supplementary Tables

**Supplementary Table 1:** Copy number alterations identified by Affymetrix 500k SNP array analysis comparing post-treatment tumor DNA to peripheral blood.

Chr	Start	End	Copy Number	SNP count	P-value	Gain/loss
chr14	105163197	106035402	1.64502	20	1.42E-08	loss
chr22	20709067	21394058	1.49146	115	0.00E+00	loss
chr12	0	133280000	2.03	NA*	NA	gain

\*Signifies that the region contains all SNPs on chromosome 12, suggesting a sub-clonal gain of this chromosome

**Supplementary Table 2:** Segments identified as copy-number gains or losses in the genome of FL patient A comparing pre-treatment tumor DNA vs peripheral blood WGSS reads.

Chr	Region start	Region end	Gain/loss	Gene(s) affected	Band(s)
chr1	103095821	103312122	gain	COL11A1	1p21.1
<b>chr1</b>	<b>105307306</b>	<b>105977510</b>	gain	RP11-414B7.1	1p21.1
chr1	121053484	121186112	loss		1p11.1, 1p11.2
<b>chr1</b>	<b>121186113</b>	<b>121186276</b>	<b>gain</b>		<b>1p11.1</b>
chr1	121186277	121186641	loss		1p11.1
<b>chr1</b>	<b>121186642</b>	<b>141477085</b>	gain		1p11.1, 1q11, 1q12
<b>chr2</b>	<b>58012705</b>	<b>58334553</b>	gain	VRK2, FANCL, AC007250.2	2p16.1
<b>chr2</b>	<b>88907314</b>	<b>89180610</b>	loss	IGK Locus	2p11.2
<b>chr2</b>	<b>89954196</b>	<b>90979647</b>	loss		2p11.2
<b>chr2</b>	<b>91631598</b>	<b>91688756</b>	loss		2p11.1
<b>chr2</b>	<b>132712328</b>	<b>132719768</b>	loss		2q21.2
<b>chr2</b>	<b>132743407</b>	<b>132746040</b>	gain		2q21.2
<b>chr2</b>	<b>193105244</b>	<b>194844048</b>	gain		2q32.3
<b>chr4</b>	<b>48786089</b>	<b>48853213</b>	loss		4p11
<b>chr4</b>	<b>49328188</b>	<b>49354810</b>	loss		4p11
<b>chr4</b>	<b>64267663</b>	<b>65106959</b>	gain	AC023156.5	4q13.1
<b>chr4</b>	<b>67946294</b>	<b>67949042</b>	loss		4q13.2
<b>chr4</b>	<b>133887173</b>	<b>134552515</b>	gain	PCDH10, AC105383.3	4q28.3
<b>chr4</b>	<b>136115277</b>	<b>137254024</b>	gain	AC104619.5, AC108867.3, AC108867.3, AC104136.5	4q28.3
<b>chr5</b>	<b>84265552</b>	<b>84734650</b>	gain	AC113412.2, AC113412.2	5q14.3
chr6	58873001	62007732	loss		6p11.1, 6q11.1
<b>chr7</b>	<b>61082565</b>	<b>61638814</b>	loss		7q11.1,

					7q11.21
<b>chr8</b>	<b>43211925</b>	<b>43217767</b>	gain		8p11.1
<b>chr8</b>	<b>43880963</b>	<b>46987966</b>	loss		8p11.1, 8q11.1
<b>chr9</b>	<b>44004560</b>	<b>44036633</b>	loss	CR848007.2, CR848007.1	9p11.2
<b>chr9</b>	<b>66533605</b>	<b>66575813</b>	loss		9q12
<b>chr10</b>	<b>38814809</b>	<b>41699890</b>	loss	AL133173.20, RP11- 453N3.5, AL133173.20, RP11-453N3.1, RP11- 96F8.1	10p11.1 ,
<b>chr10</b>	<b>41699891</b>	<b>41699983</b>	gain		10q11.1
<b>chr10</b>	<b>41699984</b>	<b>41987028</b>	loss	AL031601.3	10q11.1
<b>chr11</b>	<b>48760433</b>	<b>48856735</b>	loss		11p11.1 2, 11p11.2
chr11	51419933	51447863	loss		11p11.1 1
<b>chr12</b>	<b>34722512</b>	<b>34744090</b>	loss		12p11.1
<b>chr13</b>	<b>61882511</b>	<b>62676590</b>	gain		13q21.3 1
<b>chr13</b>	<b>76560040</b>	<b>76760141</b>	gain	MYCBP2	13q22.3
chr14	18070001	18109142	loss		14q11.1
<b>chr14</b>	<b>105278014</b>	<b>106038921</b>	loss	IGH locus	14q32.3 3
<b>chr16</b>	<b>33775618</b>	<b>33809954</b>	loss		16p11.2
<b>chr16</b>	<b>33870617</b>	<b>34053870</b>	loss	AC136932.4, AC136932.4	16p11.2
chr16	35072626	45016121	loss		16p11.1 ,
chr17	1	242951149	gain		16q11.1 ,
<b>chr18</b>	<b>96814</b>	<b>101400</b>	loss	AP001005.6	18p11.3 2
<b>chr18</b>	<b>16764927</b>	<b>16774290</b>	loss		18q11.1
chr19	32423734	32423881	gain		19q12
chr19	32423882	32445251	loss		19q12
<b>chr21</b>	<b>9736820</b>	<b>9745481</b>	loss		21p11.2
<b>chr21</b>	<b>9834768</b>	<b>9878713</b>	loss		21p11.2
<b>chr22</b>	<b>15228227</b>	<b>15242356</b>	loss		22q11.1
<b>chr22</b>	<b>20707883</b>	<b>21574967</b>	loss	IGL locus	22q11.2 2
chrX	35787300	36284003	gain	CXorf22, AL590065.3, CXorf59, RP13- 172P16.1, AL606467.5, CXorf30	Xp21.1
chrX	58577552	61768896	loss		Xp11.1, Xq11.1

This table summarizes the data displayed in Supplementary Figure 6. Notably, very few genes are in the regions denoted as gains or losses. The three major losses identified here (i.e. affecting a large number of genes) affect the immunoglobulin gene clusters where deletions are expected to occur as a normal part of immunoglobulin class switching. The events indicated in bold overlap to a large extent with known germline variations (as determined by comparison with the database of genomic variants).

**Supplementary Table 3:** Novel coding SNVs (candidate mutations) identified in WTSS.

Position	Gene	Ref base	Non-ref Base	WTSS Reads (ref allele)	WTSS Reads (non-ref allele)	Ref Amino Acid	Non-ref Amino Acid (or * for STOP)
chr1:1131748	TNFRSF18	T	C	2	3	S	G
chr1:10381723	PGD	C	T	30	22	R	C
chr1:20315465	PLA2G2D	C	T	70	37	G	S
chr1:21762347	ALPL	G	A	11	12	R	H
chr1:43670083	KIAA0467	G	A	1	5	R	H
chr1:46541451	LRRC41	G	C	2	4	P	R
chr1:89296499	GBP1	C	T	14	12	S	N
chr1:148994106	CTSS	A	C	217	48	Y	D
chr1:149503078	PSMD4	G	T	0	2	V	F
chr1:156331454	KIRREL	A	C	5	3	T	P
chr1:158759594	SLAMF6	A	C	20	7	F	V
chr1:172106402	ZBTB37	G	T	4	2	S	I
chr2:11401593	ROCK2	T	C	3	5	I	V
chr2:27135770	AGBL5	G	A	0	2	V	I
chr2:96237694	STARD7	G	A	18	32	P	L
chr2:108714301	RANBP2	A	G	18	8	K	R
chr2:201845626	CASP8	A	G	17	12	I	M
chr2:203470334	WDR12	C	T	4	4	G	R
chr2:203819783	CYP20A1	G	T	14	5	L	F
chr2:231371767	CAB39	A	G	62	58	I	V
chr3:9686176	MTMR14	G	A	44	13	R	Q
chr3:69313282	FRMD4B	T	G	7	4	N	T
chr3:109885838	DZIP3	C	T	5	9	P	L
chr3:122899503	GOLGB1	G	A	2	8	L	F
chr3:126435054	SLC12A8	T	G	11	3	K	N
chr3:185553595	CLCN2	C	T	0	2	R	Q
chr3:197872542	LRRC33	C	T	6	9	S	L
chr4:357199	ZNF141	A	T	2	2	T	S
chr4:1813073	LETM1	G	T	1	4	N	K
chr4:8284650	SH3TC1	A	G	6	3	M	V
chr4:10054305	ZNF518B	T	C	1	4	I	V
chr4:20228276	SLIT2	C	T	4	3	A	V
chr4:71131547	C4orf7	T	G	76	13	V	G
chr4:78090084	SEPT11	G	C	4	5	A	P
chr4:164726502	MARCH1	C	T	53	62	S	N
chr5:102460210	GIN1	C	A	2	2	V	F
chr6:30787176	MDC1	T	C	5	5	Q	R
chr6:32597800	HLA-DRB5	C	T	1	4	R	Q
chr7:5319475	TNRC18	G	C	10	5	L	V

chr7:5393918	TNRC18	A	C	11	3	V	G
chr7:24725289	DFNA5	G	A	10	5	Q	*
chr7:66120549	TYW1	A	G	4	6	H	R
chr7:101885258	ALKBH4	A	C	2	4	I	S
chr7:101900231	LRWD1	C	T	9	8	P	L
chr7:107942613	PNPLA8	G	A	8	12	R	C
chr7:124279208	POT1	C	A	4	8	Q	H
chr7:130778829	MKLN1	T	G	14	7	I	S
chr7:139825570	MKRN1	C	A	6	4	G	V
chr7:148139661	EZH2	A	G	11	4	Y	H
chr8:19732261	INTS10	A	C	32	10	E	A
chr8:145592397	CPSF1	G	A	8	9	R	C
chr9:116826116	TNC	C	T	69	59	R	H
chr9:133324433	KIAA0515	G	A	23	11	A	T
chr9:134207924	SETX	A	C	14	12	L	V
chr9:138396112	SNAPC4	C	T	6	6	V	I
chr9:139281277	COBRA1	A	G	20	29	T	A
chr9:139569822	WDR85	C	T	9	3	R	Q
chr10:45565563	FAM21D	T	C	3	4	M	T
chr10:71575478	TYSND1	C	A	6	9	V	L
chr10:75205265	FUT11	G	C	0	2	A	P
chr10:79465368	RPS24	A	C	67	116	I	L
chr10:89464598	PAPSS2	C	G	13	11	Q	E
chr10:90764005	FAS	C	T	46	9	Q	*
chr10:104668340	CNNM2	G	A	1	3	R	Q
chr11:46787573	CKAP5	C	T	13	7	A	T
chr11:58736424	MPEG1	G	A	28	36	T	I
chr11:61320311	FEN1	C	G	11	6	P	R
chr11:65113802	EHBP1L1	G	A	8	6	G	D
chr11:65142149	PCNXL3	A	G	18	15	Q	R
chr11:67021989	PITPNM1	C	T	28	25	R	Q
chr11:73446107	C2CD3	C	A	7	3	Q	H
chr11:114880310	CADM1	C	G	0	22	V	L
chr11:118270511	BLR1	C	T	115	54	Q	*
chr12:2668007	CACNA1C	G	A	1	2	R	Q
chr12:29556302	TMTCT1	G	C	1	2	Q	E
chr12:42428601	PUS7L	G	A	4	5	S	L
chr12:46429582	RAPGEF3	A	G	0	3	L	P
chr12:47717214	MLL2	T	C	10	6	M	V
chr12:55785346	STAT6	G	T	77	45	Q	K
chr12:93138029	PLXNC1	G	A	19	19	R	Q
chr12:107707010	SSH1	G	A	8	7	T	I
chr13:72267613	C13orf24	C	T	6	6	R	C
chr14:20012815	NP	C	T	35	19	T	I
chr14:72823631	NUMB	T	G	30	6	T	P
chr14:73109889	ACOT2	G	A	0	2	G	S
chr14:87486760	GALC	G	A	12	8	S	F

chr15:38856063	DNAJC17	T	G	3	2	E	A
chr15:41604904	MAP1A	A	G	2	2	E	G
chr16:2203837	PGP	C	T	6	9	G	R
chr16:3726768	CREBBP	A	G	33	22	Y	H
chr16:30640059	SRCAP	C	A	16	19	T	K
chr16:68336417	NOB1	T	C	5	12	T	A
chr17:4391586	MYBBP1A	G	A	10	13	A	V
chr17:8082622	C17orf68	C	G	22	20	S	T
chr17:25730809	CPD	C	G	4	2	P	A
chr17:34125400	MLLT6	T	A	48	25	S	T
chr17:37167297	JUP	C	T	11	92	V	I
chr17:71568013	SRP68	T	C	19	25	T	A
chr18:19738538	LAMA3	A	G	3	2	D	G
chr18:31960233	SLC39A6	C	T	16	24	E	K
chr19:5874888	RANBP3	T	C	11	11	N	S
chr19:19121087	MEF2B	T	C	56	45	Y	C
chr19:42827422	ZFP30	C	A	5	5	C	F
chr19:50267501	ZNF342	A	C	8	4	V	G
chr19:55084855	IL4I1	C	T	5	5	V	M
chr20:23013733	CD93	C	T	5	9	R	H
chr20:60338313	LAMA5	A	C	2	2	V	G
chr21:14675411	STCH	C	A	12	6	G	V
chr21:41733614	MX1	T	G	20	5	V	G
chr21:46406377	C21orf56	C	T	4	2	A	T
chr22:31585244	TIMP3	C	T	64	68	R	W
chrX:18846780	PHKA2	T	C	8	6	I	V
chrX:47803931	ZNF630	G	C	0	2	Q	E
chrX:48647000	SLC35A2	C	A	4	5	R	L
chrX:53669571	HUWE1	G	T	0	2	P	T
chrX:122694531	THOC2	A	C	9	3	V	G
chrX:148500835	TMEM185A	T	C	1	2	K	E
chrX:148500838	TMEM185A	A	C	2	2	W	G
chrX:150324177	VMA21	A	G	7	12	K	E
chrX:152873134	HCFC1	T	G	27	12	T	P

This table includes the novel variants identified (by SNVMix) in the FL patient A WTSS library HS0804 and confirmed by at least 1 high-quality, unambiguously mapped read in the tumor WGSS library. “ref” refers to a match to the reference genome and “non-ref” refers to an alternate nucleotide or amino acid. Events also present in the germline WGSS library have been removed. Due to low coverage in the germline genome, this table likely contains many remaining un-recognized germline variants along with artifacts from sequencing and alignment. The number of high quality reads supporting each allele are included as coverage can be used to further refine this list.

**Supplementary Table 4:** Coverage of *EZH2* and exon 15 in all RNA-seq libraries.

Library/sample	Type	Subtype	<i>EZH2</i> depth	Reads supporting wild-type Y641	Reads supporting a Y641 mutation
HS0639*	DLBCL	GCB	44.18	31	51
HS0640*	DLBCL	GCB	54.88	29	19
HS0644	DLBCL	GCB	40.01	29	0
HS0645	DLBCL	U	71.62	76	0
HS0647	DLBCL	GCB	46.51	69	0
HS0648*	DLBCL	GCB	69.80	53	56
HS0649	DLBCL	ABC	61.81	88	0
HS0650	DLBCL	GCB	25.58	45	0
HS0651	DLBCL	GCB	43.25	56	0
HS0652	DLBCL	ABC	47.23	62	1
HS0653	DLBCL	GCB	42.38	74	0
HS0656	DLBCL	U	50.29	50	0
HS0669	Centroblasts	NA	72.70	53	1
HS0670	Centroblasts	NA	122.81	107	1
HS0747	DLBCL	U	7.56	10	0
HS0748	DLBCL	ABC	177.54	294	1
HS0749	DLBCL	GCB	85.98	70	1
HS0750	DLBCL	ABC	37.22	39	1
HS0751	DLBCL	ABC	34.17	35	0
HS0804*	FL	NA	13.61	11	4
HS0926	DLBCL	U	9.98	24	0
HS0927	DLBCL	U	88.68	24	0
HS0928	DLBCL	ABC	186.93	40	0
HS0929	DLBCL	GCB	90.14	25	0
HS0930	DLBCL	GCB	54.89	8	0
HS0931	DLBCL	U	103.87	14	0
HS0932	DLBCL	ABC	62.27	82	0
HS0933	DLBCL	GCB	78.83	89	1
HS0934	DLBCL	U	27.65	44	0
HS0935	DLBCL	U	71.63	111	0
HS0940	DLBCL	ABC	17.02	27	0
HS0941	DLBCL	U	11.71	20	0
HS0942*	DLBCL	GCB	5.38	1	4
HS0943	DLBCL	U	22.38	26	0

\*indicates a library generated from a patient sample with the *EZH2* Y641 mutation

**Supplementary Table 5:** All *EZH2* mutants detected by Sanger sequencing in FL and DLBCL.

ID	Diagnosis	<i>EZH2</i> mutation	Position (chr7)	Nucleotide change	Amino Acid change
00-13940	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
00-15694	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
00-19845	DLBCL (GCB)	Y641 MUT	148139660	A->C	Y->S
00-22287	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
01-19969	DLBCL (GCB)	Y641 MUT	148139660	A->C	Y->S
02-26353	DLBCL (GCB)	Y641 MUT	148139660	A->C	Y->S
02-30647	DLBCL (GCB)	Y641 MUT	148139661	T->C	Y->H
04-11156	DLBCL (GCB)	Y641 MUT	148139660	A->C	Y->S
04-28216	DLBCL (GCB)	Y641 MUT	148139660	A->C	Y->S
04-39242	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
05-11328	DLBCL (GCB)	Y641 MUT	148139661	T->C	Y->H
05-23110	DLBCL (GCB)	Y641 MUT	148139661	T->C	Y->H
05-25439	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
05-32947	DLBCL (GCB)	Y641 MUT	148139661	T->C	Y->H
06-24718	DLBCL (GCB)	Y641 MUT	148139660	A->C	Y->S
07-30109	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
07-30628	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
98-14032	DLBCL (GCB)	Y641 MUT	148139660	A->T	Y->F
01-15178	DLBCL (NA)	Y641 MUT	148139661	T->C	Y->H
01-24152	DLBCL (NA)	Y641 MUT	148139660	A->C	Y->S
01-28389	DLBCL (NA)	Y641 MUT	148139660	A->T	Y->F
02-24981	DLBCL (NA)	Y641 MUT	148139661	T->A	Y->N
03-11110	DLBCL (NA)	Y641 MUT	148139660	A->T	Y->F
03-28045	DLBCL (NA)	Y641 MUT	148139660	A->C	Y->S
05-12131	DLBCL (NA)	Y641 MUT	148139660	A->T	Y->F
05-26898	DLBCL (NA)	Y641 MUT	148139661	T->C	Y->H
06-27034	DLBCL (NA)	Y641 MUT	148139660	A->T	Y->F
96-20883	DLBCL (NA)	Y641 MUT	148139661	T->A	Y->N
99-22226	DLBCL (NA)	Y641 MUT	148139660	A->T	Y->F
99-29859	DLBCL (PMBCL)	Y641 MUT	148139660	A->T	Y->F
01-11023	FL Grade 1	Y641 MUT	148139660	A->T	Y->F
02-23246	FL Grade 1	Y641 MUT	148139661	T->A	Y->N
05-12472	FL Grade 1	Y641 MUT	148139661	T->A	Y->N
06-12968	FL Grade 1	Y641 MUT	148139661	T->C	Y->H
06-19522	FL Grade 1	Y641 MUT	148139660	A->T	Y->F
06-19817	FL Grade 1	Y641 MUT	148139660	A->T	Y->F
89-33903	FL Grade 1	Y641 MUT	148139660	A->T	Y->F
89-37479	FL Grade 1	Y641 MUT	148139660	A->T	Y->F
90-34286	FL Grade 1	Y641 MUT	148139660	A->T	Y->F
95-13715	FL Grade 1	Y641 MUT	148139660	A->C	Y->S

99-17919* <sup>+</sup>	FL Grade 1, DLBCL	WT in FL, Y641 MUT in DLBCL	148139661	T->A	Y->N
99-30068*	FL Grade 1, DLBCL	WT in FL, Y641 MUT in DLBCL	148139661	T->A	Y->N
06-16058	FL Grade 2	Y641 MUT	148139660	A->T	Y->F
06-23851	FL Grade 2	Y641 MUT	148139660	A->T	Y->F
06-30133	FL Grade 2	Y641 MUT	148139660	A->T	Y->F
96-26853	FL Grade 2	Y641 MUT	148139661	T->A	Y->N
96-26853	FL Grade 2	SET MUT	148139677	T->A	N->K
00-27081*	FL Grade 2, DLBCL	WT in FL, Y641 MUT in DLBCL	148139660	A->G	Y->C
03-11874*	FL Grade 2, DLBCL	Y641 MUT in FL, WT in DLBCL	148139660	A->T	Y->F
95-24059*	FL Grade 2, DLBCL	Y641 MUT in FL, WT in DLBCL	148139660	A->T	Y->F
94-12812	FL Grade 3	Y641 MUT	148139660	A->T	Y->F
04-40070	FL Grade 3A	Y641 MUT	148139660	A->C	Y->S
02-18484* <sup>+</sup>	FL Grade 3A, DLBCL	WT in FL, Y641 MUT in DLBCL	148139660	A->C	Y->S

\*FL patients with paired samples pre- and post-transformation to DLBCL

<sup>+</sup>Visual assessment of Sanger trace files revealed evidence for the same Y641 mutation in the matched sample originally deemed wild type by automated analysis of Sanger sequence data (peak height was below threshold)

**Supplementary Table 6:** Ultra deep targeted re-sequencing of normal B-cell populations and non-GCB lymphomas.

Index*	Sample ID	Type	Maximum high-quality matches at either non-synonymous site (codon 641)	Maximum high-quality mismatches at either non-synonymous site (codon 641)	Maximum % high-quality mismatches at either non-synonymous site (codon 641)	amplicon average % high-quality mismatch
ACGATA	V00196	FL	2430	207	7.850	0.145
GTAGAG	4299	FL	2749	194	6.592	0.144
TGCTGG	9425	FL	1583	263	14.24	0.079
ATCACG	V00180	Tonsil	1276	0	0.000	0.084
GATCAG	V00522	Tonsil	2073	2	0.112	0.082
AACCCC	V00523	Tonsil	2697	4	0.158	0.105
ACCCAG	V00524	Tonsil	3049	0	0.000	0.090
AGCGCT	V00525	Tonsil	1501	0	0.000	0.076
CAAAAG	V00526	Tonsil	5585	3	0.061	0.117
CCAACA	V00527	Tonsil	3514	6	0.182	0.235
CTAGCT	V00528	Tonsil	1508	1	0.066	0.088
GATGCT	V00530	Tonsil	1182	2	0.189	0.100
TAATCG	V00531	Tonsil	1764	0	0.000	0.092
TGAATG	V00537	Tonsil	2610	2	0.084	0.112
AGTTCC	V00538	Tonsil	1331	1	0.075	0.074
CGATGT	V00539	Tonsil	1326	1	0.085	0.071
TAGCTT	V00540	Tonsil	1906	1	0.060	0.076
AACTTG	V00541	Tonsil	8556	7	0.101	0.264
ACCGGC	V00542	Tonsil	1657	2	0.139	0.086
AGGCCG	V00543	Tonsil	5545	3	0.061	0.386
CAACTA	V00545	Tonsil	2101	2	0.095	0.236
CCACGC	V00546	Tonsil	1625	0	0.000	0.075
CTATAC	V00548	Tonsil	4836	4	0.083	0.252
GCAAGG	V00550	Tonsil	1706	2	0.117	0.087
TACAGC	V00553	Tonsil	2258	1	0.049	0.134
TGCCAT	V00557	Tonsil	1558	1	0.071	0.082
ATGTCA	V00563	SLL	1327	0	0.000	0.078
TTAGGC	V00573	SLL	1705	3	0.194	0.098
GGCTAC	V00595	SLL	2222	2	0.104	0.074
AAGACT	V00600	SLL	10225	4	0.044	0.104
ACGATA	V00609	MCLN	1444	1	0.077	0.200
ATAATT	V00611	MCLD	18089	11	0.070	0.346
CACCGG	V00625	MCLN	5010	3	0.069	0.352
CCCATG	V00641	SLL	2961	2	0.072	0.110
CTCAGA	V00646	SLL	18903	8	0.046	0.096
GCACTT	V00649	SLL	1753	1	0.064	0.231
TATAAT	V00654	MCLN	105394	45	0.046	0.304

TGCTGG	V00655	MCLN	2805	2	0.079	0.101
CCGTCC	V00660	MCLD	1614	2	0.124	0.094
TGACCA	V00663	SLL	1988	1	0.057	0.160
CTTGTA	V00668	SLL	7808	8	0.113	0.432
AAGCGA	V00673	MCLD	3535	3	0.094	0.100
ACTCTC	V00677	SLL	1769	1	0.056	0.094
ATACGG	V00695	MCLD	3276	3	0.099	0.090
CACGAT	V00701	SLL	1673	1	0.065	0.108
CCCCCT	V00727	MCLD	133041	73	0.059	0.152
CTGCTG	V00732	MCLD	30803	20	0.075	0.133
GCCGCG	V00734	SLL	2329	1	0.048	0.081
TCATTC	V00735	MCLD	2055	1	0.055	0.291
TGGCGC	V00737	MCLN	8485	4	0.051	0.135
GTAGAG	V00738	SLL	1305	1	0.085	0.157
ACAGTG	V00739	MCLD	1305	1	0.077	0.070
AAACAT	V00741	MCLD	1298	1	0.083	0.079
AAGGAC	V00744	MCLN	9418	3	0.040	0.178
ACTGAT	V00750	MCLN	1717	2	0.131	0.248
ATCCTA	V00752	MCLD	2267	1	0.049	0.128
CACTCA	V00762	MCLD	29623	18	0.067	0.334
CCGCAA	V00764	MCLD	1197	1	0.094	0.161
GAAACC	V00769	SLL	14238	8	0.059	0.298
GCCTTA	V00770	SLL	3248	1	0.033	0.104
TCCCGA	V00772	SLL	3872	4	0.115	0.109
TTCGAA	V00774	MCLD	23240	10	0.046	0.124
GTCCGC	V00775	SLL	1330	0	0.000	0.075
GCCAAT	V00779	SLL	2163	1	0.051	0.093
AAAGCA	V00781	SLL	1571	2	0.141	0.235
AATAGG	V00786	SLL	15048	9	0.062	0.073
AGAAGA	V00789	MCLD	40993	20	0.050	0.201
ATCTAT	V00790	SLL	37055	15	0.042	0.216
CAGGCG	V00808	MCLD	12481	6	0.052	0.075
CCTTAG	V00809	SLL	88603	49	0.061	0.934
GAATAA	V00810	SLL	61609	23	0.039	0.265
GCTCCA	V00814	SLL	75635	52	0.069	0.262
TCGAAG	V00819	SLL	69900	21	0.033	0.200
TTCTCC	V00826	SLL	19348	7	0.043	0.111
GTGAAA	V00830	SLL	4003	4	0.108	0.117
CAGATC	V00831	SLL	1484	2	0.152	0.086
AAATGC	V00832	MCLN	1662	0	0.000	0.080
ACAAAC	V00837	SLL	24	0	0.000	0.041
AGATAG	V00838	SLL	1583	1	0.063	0.094
ATGAGC	V00929	MCL	10327	15	0.155	0.459
CATGGC	V00930	MCL	3064	1	0.036	0.104
CGAGAA	V00946	PTCL	13428	8	0.064	0.073
GACGGA	V00951	PTCL	1288	2	0.169	0.087
GGCACCA	V00956	PTCL	2043	2	0.098	0.292

TCGGCA	V00957	PTCL	2694	3	0.121	0.105
AGGTTT	V00959	PTCL	4243	3	0.078	0.138
GTGGCC	V00960	PTCL	1586	0	0.000	0.095
ACTTGA	V00961	PTCL	4625	2	0.047	0.113
AACAAA	V00963	PTCL	3098	2	0.065	0.097
ACATCT	V00964	PTCL	428478	220	0.051	1.028
AGCATC	V00965	PTCL	111	1	0.952	0.115
ATT CCT	V00967	PTCL	17	0	0.000	0.059
CAT TTT	V00970	PTCL	74	0	0.000	0.125
CGGAAT	V00971	PTCL	116	0	0.000	0.104
GATATA	V00975	PTCL	22105	8	0.038	0.142
GGCCTG	V00977	PTCL	11	0	0.000	0.000
TCTACC	V00980	PTCL	10	0	0.000	0.207
AGTCAA	V00984	PTCL	219	0	0.000	0.102
ATCACG	V00989	PTCL	2453	1	0.042	0.073
GATCAG	V00991	PTCL	2667	1	0.039	0.058
AACCCC	V00993	PTCL	1511	1	0.068	0.052
ACCCAG	V00996	PTCL	1953	1	0.051	0.058
AGCGCT	V00997	PTCL	2268	2	0.088	0.225
CAAAAG	V00999	PTCL	3729	2	0.054	0.061
CCAACA	V01000	PTCL	1592	1	0.063	0.112
CTAGCT	V01002	PTCL	29436	12	0.042	0.246
		CD77				
GATGCT	21148-CB	+ CB	2179	1	0.047	0.153
TAATCG	2407-CB	CD77 + CB	3563	0	0.000	0.071
TGAATG	2424-CB	CD77 + CB	4644	2	0.044	0.067
AGTTCC	12307-CB	CD77 + CB	8728	2	0.023	0.073
CGATGT	3412-CB	CD77 + CB	1583	0	0.000	0.052
TAGCTT	4915-CB	CD77 + CB	2143	0	0.000	0.073
AACTTG	4920-CB	CD77 + CB	1982	1	0.052	0.051
ACCGGC	120808-CB	CD77 + CB	1642	2	0.122	0.191

\*Some indexes are duplicates because not all samples were sequenced in the same library.

Samples indicated in red are positive controls demonstrating efficient recovery of known Y641 mutations in FL samples

This table summarizes the result from ultra deep targeted re-sequencing the genomic region flanking codon 641 in *EZH2*. Note that different index sequences result in variations in net sequence coverage. Shown are three FL samples (red) with known Y641 mutations. Of the two non-synonymous sites in codon 641, the site with the larger number of high-quality mismatches (>Q20) was chosen. The first numeric column states the total number of high-quality bases matching the reference sequence at that position.

The next column states the total number of high-quality mismatches at this position. The next column is these two values represented as a percentage (to compensate for variable coverage between amplicons). The rightmost column states the average of this percentage across all positions in the amplicon. Clearly, known mutants show a strong enrichment of high-quality mismatches at this codon. In a pure tumor sample, this value would be expected to approach 50% in a heterozygous mutant.

**Supplementary Table 7:** Primer sequences.

Primer name	Sequence
EZH2_015R3	TCTCAGCAGCTTCACGTTG
EZH2_015F	CAGGTTATCAGTGCCTTACCTCTCC
Multiplexing Adapter 1	GATCGGAAGAGCACACGTCT
Multiplexing Adapter 2	ACACTTTCCCTACACGCTCTCCGATCT
Primer 1.0	AATGATA CGGCGACCACCGAGATCTACACTTTCCCTACACGACGCTCTCCGATCT
Primer 2.0	GTGACTGGAGTT CAGACGTGTGCTCTCCGATCT
EZH2 ASP_1	TGTAAAACGACGCCAGTCTGGACTACAAGTATGCACCAACC
EZH2 ASP_2	CAGGAAACAGCTATGACACCAACACCACCAAAAGGTTTCT